

Filtering and Smoothing for Linear Discrete-Time Distributed Parameter Systems Based on Wiener-Hopf Theory with Application to Estimation of Air Pollution

SIGERU OMATU, MEMBER, IEEE, AND JOHN H. SEINFELD

Abstract—Optimal filtering and smoothing algorithms for linear discrete-time distributed parameter systems are derived by a unified approach based on the Wiener-Hopf theory. The Wiener-Hopf equation for the estimation problems is derived using the least-squares estimation error criterion. Using the basic equation, three types of the optimal smoothing estimators are derived, namely, fixed-point, fixed-interval, and fixed-lag smoothers. Finally, the results obtained are applied to estimation of atmospheric sulfur dioxide concentrations in the Tokushima prefecture of Japan.

I. INTRODUCTION

A NUMBER of important physical phenomena may be modeled as discrete-time distributed parameter systems. When estimation problems are encountered in such systems, the measurements are also frequently discrete in time. A great deal of work has been carried out on estimation problems for continuous-time distributed parameter systems [1]–[4]. Tzafestas [5], [6] and Nagamine *et al.* [7] have derived optimal estimators for discrete-time distributed parameter systems. Tzafestas employed a Bayesian approach, where Nagamine *et al.* considered only the filtering problem based on the Wiener-Hopf theory. Recently, Bencala and Seinfeld [3] have derived the optimal filter for

continuous-time distributed parameter systems with discrete-time observations by the Wiener-Hopf approach.

The object of this paper is twofold. First, we seek to derive optimal filtering and smoothing algorithms for discrete-time distributed parameter systems by a unified Wiener-Hopf approach. Fixed-point, fixed-interval, and fixed lag smoothers are considered. Second, we wish to apply the results to the estimation of atmospheric sulfur dioxide concentrations in the Tokushima prefecture of Japan.

II. DESCRIPTION OF THE DISTRIBUTED PARAMETER SYSTEM

Let D be a bounded open domain of an r -dimensional Euclidean space with smooth boundary ∂D . The spatial coordinate vector will be denoted by $x = (x_1, \dots, x_r) \in D$. Consider a linear distributed parameter system described by

$$u(k+1, x) = \mathcal{L}_x u(k, x) + G(k, x)w(k, x), \quad x \in D \quad (1)$$

where $u(k+1, x)$ is an n -dimensional vector function of the system, $w(k, x)$ is a vector-valued Gaussian process, \mathcal{L}_x is a linear spatial matrix differential operator, and $G(k, x)$ is a known matrix function.

The initial and boundary conditions are given by

$$u(0, x) = u_0(x) \quad (2)$$

$$\Gamma_\xi u(k+1, \xi) = S(k+1, \xi), \quad \xi \in \partial D \quad (3)$$

$$\Gamma_\xi [\cdot] = \alpha(\xi)[\cdot] + (1 - \alpha(\xi))\partial[\cdot]/\partial n \quad (4)$$

Manuscript received February 3, 1981; revised August 3, 1981. This work was supported by NASA Research Grant NAG1-71.

S. Omatu was with the Department of Chemical Engineering, California Institute of Technology, Pasadena, CA, on leave from the Department of Information Science and Systems Engineering, University of Tokushima, Tokushima, Japan.

J. H. Seinfeld is with the Department of Chemical Engineering, California Institute of Technology, Pasadena, CA 91125.

where n is an exterior normal vector to the boundary ∂D at a point $\xi \in \partial D$ and $\alpha(\xi)$ is a function of class C^2 on ∂D satisfying $0 \leq \alpha(\xi) \leq 1$. $S(k+1, \xi)$ denotes a source function at the boundary and is assumed to be known.

Assume that $u_0(x)$ is a Gaussian random function the mean and covariance functions of which are given by

$$E[u_0(x)] = 0 \quad (5)$$

$$E[u_0(x)u_0'(y)] = P_0(x, y) \quad (6)$$

where $E[\cdot]$ and the prime symbol denote the expectation and transpose operators, respectively.

Let the observed data be taken at m points, $x^1, \dots, x^m \in \bar{D} = D \cup \partial D$ and let an mn -dimensional column vector $u_m(k)$ be defined by

$$u_m(k) = \text{Col}[u(k, x^1), \dots, u(k, x^m)]. \quad (7)$$

Let the observations be related to the states by

$$z(k) = H(k)u_m(k) + v(k), \quad (8)$$

where $z(k)$ is a p -dimensional observations vector at the m observation points, $x^1, \dots, x^m \in \bar{D}$, $H(k)$ is a known $p \times mn$ matrix, and $v(k)$ is a p -dimensional vector-valued white Gaussian process. Assume that the white Gaussian process $w(k, x)$ in (1) and $v(k)$ in (8) are statistically independent of each other and also independent of the initial condition $u_0(x)$. Their mean and covariance functions are given by

$$E[w(k, x)] = 0, \quad E[v(k)] = 0 \quad (9)$$

$$E[w(k, x)w'(s, y)] = Q(k, x, y)\delta_{ks}, \quad x, y \in D \quad (10)$$

$$E[v(k)v'(s)] = R(k)\delta_{ks}, \quad (11)$$

where δ_{ks} is the Kronecker delta function, and $Q(k, x, y)$ and $R(k)$ are symmetric positive-semidefinite and positive-definite matrices, respectively.

III. DESCRIPTION OF THE ESTIMATION PROBLEMS

The general problem considered here is to find an estimate $\hat{u}(\tau, x/k)$ of the state $u(\tau, x)$ at time τ based on the measurement data z_0^k , denoting a family of $z(\sigma)$ from $\sigma = 0$ up to the present time k . Specifically, for $\tau > k$ we have the prediction problem, for $\tau = k$ the filtering problem, and for $\tau < k$ the smoothing problem. As in the Kalman-Bucy approach, an estimate $\hat{u}(\tau, x/k)$ of $u(\tau, x)$ is sought through a linear operation on the past and present observation values z_0^k as follows:

$$\hat{u}(\tau, x/k) = \sum_{\sigma=0}^k \tilde{F}(\tau, x, \sigma)z(\sigma) \quad (12)$$

where $\tilde{F}(\tau, x, \sigma)$ is an $n \times p$ matrix kernel function.

To differentiate between the prediction, filtering, and smoothing problems, we replace (12) with different notation for each problem:

1) Prediction ($\tau > k$)

$$\hat{u}(\tau, x/k) = \sum_{\sigma=0}^k A(\tau, x, \sigma)z(\sigma). \quad (13)$$

2) Filtering ($\tau = k$)

$$\hat{u}(k, x/k) = \sum_{\sigma=0}^k F(\tau, x, \sigma)z(\sigma). \quad (14)$$

3) Smoothing ($\tau < k$)

$$\hat{u}(\tau, x/k) = \sum_{\sigma=0}^k B(\tau, k, x, \sigma)z(\sigma). \quad (15)$$

The estimation error is denoted by $\tilde{u}(\tau, x/k)$,

$$\tilde{u}(\tau, x/k) = u(\tau, x) - \hat{u}(\tau, x/k). \quad (16)$$

The estimate $\hat{u}(\tau, x/k)$ that minimizes

$$J(\hat{u}) = E[\|\tilde{u}(\tau, x/k)\|^2] \quad (17)$$

is said to be optimal, where $\|\cdot\|$ denotes the Euclidian norm.

Theorem 1: (Wiener-Hopf theory). A necessary and sufficient condition for the estimate $\hat{u}(\tau, x/k)$ to be optimal is that the following Wiener-Hopf equation holds for $\alpha = 0, 1, \dots, k$ and $x \in \bar{D}$,

$$\sum_{\sigma=0}^k \tilde{F}(\tau, x, \sigma)E[z(\sigma)z'(\alpha)] = E[u(\tau, x)z'(\alpha)]. \quad (18)$$

Furthermore, (18) is equivalent to

$$E[\tilde{u}(\tau, x/k)z'(\alpha)] = 0 \quad (19)$$

for $\alpha = 0, 1, \dots, k$ and $x \in \bar{D}$.

Proof: Let $F_{\Delta}(\tau, x, \sigma)$ be an $n \times p$ matrix function and let ϵ be a scalar-valued parameter. The trace of the covariance of the estimate,

$$\hat{u}_{\epsilon}(\tau, x/k) = \sum_{\sigma=0}^k (\tilde{F}(\tau, x, \sigma) + \epsilon F_{\Delta}(\tau, x, \sigma))z(\sigma)$$

is given by

$$\begin{aligned} J(\hat{u}_{\epsilon}) = E \left[\left\| u(\tau, x) - \hat{u}_{\epsilon}(\tau, x/k) \right\|^2 \right] \\ = E \left[\left\| u(\tau, x) - \sum_{\sigma=0}^k \tilde{F}(\tau, x, \sigma)z(\sigma) - \epsilon \sum_{\sigma=0}^k F_{\Delta}(\tau, x, \sigma)z(\sigma) \right\|^2 \right] \\ = E \left[\left\| u(\tau, x) - \sum_{\sigma=0}^k \tilde{F}(\tau, x, \sigma)z(\sigma) \right\|^2 \right] \\ - 2\epsilon E \left[\tilde{u}'(\tau, x/k) \sum_{\sigma=0}^k F_{\Delta}(\tau, x, \sigma)z(\sigma) \right] \\ + \epsilon^2 E \left[\left\| \sum_{\sigma=0}^k F_{\Delta}(\tau, x, \sigma)z(\sigma) \right\|^2 \right]. \end{aligned}$$

A necessary and sufficient condition for $\hat{u}(\tau, x/k)$ to be optimal is that

$$\left. \frac{\partial J(\hat{u}_{\epsilon})}{\partial \epsilon} \right|_{\epsilon=0} = 0,$$

that is,

$$E \left[\tilde{u}'(\tau, x/k) \sum_{\sigma=0}^k F_{\Delta}(\tau, x, \sigma)z(\sigma) \right] = 0$$

for any $n \times p$ matrix $F_\Delta(\tau, x, \sigma)$. Using the relation between the trace and inner product yields

$$\begin{aligned} & E \left[\tilde{u}'(\tau, x/k) \sum_{\sigma=0}^k F_\Delta(\tau, x, \sigma) z(\sigma) \right] \\ &= \text{tr} \left[E \left[\tilde{u}(\tau, x/k) \sum_{\sigma=0}^k z'(\sigma) F_\Delta'(\tau, x, \sigma) \right] \right] \\ &= \sum_{\sigma=0}^k \text{tr} [E[\tilde{u}(\tau, x/k) z'(\sigma)] F_\Delta'(\tau, x, \sigma)] = 0. \end{aligned}$$

Setting $F_\Delta(\tau, x, k) = E[\tilde{u}(\tau, x/k) z'(\sigma)]$ in the above equation, it follows that (19) is a necessary condition for $\hat{u}(\tau, x/k)$ to be optimal. Sufficiency of (19) also follows from the above equation. Q.E.D.

Corollary 1: (Orthogonal projection lemma). The following orthogonality condition holds,

$$E[\tilde{u}(\tau, x/k) \hat{u}'(\xi, y/k)] = 0, \quad x, y \in \bar{D} \quad (20)$$

where ξ is any time instant, for example, $\xi < k$, $\xi = k$ or $\xi > k$.

Proof: Multiplying each side of (19) by $\tilde{F}(\xi, y, \alpha)$ and summing from $\sigma = 0$ to $\sigma = k$ yields

$$E \left[\tilde{u}(\tau, x/k) \sum_{\alpha=0}^k z'(\alpha) \tilde{F}'(\xi, y, \alpha) \right] = 0.$$

Substituting (12) into the above equation yields (20).

Q.E.D.

Then the following lemma can be proved.

Lemma 1: (Uniqueness of the optimal kernel). Let $\tilde{F}(\tau, x, \sigma)$ and $\tilde{F}(\tau, x, \sigma) + N(\tau, x, \sigma)$ be optimal matrix kernel functions satisfying the Wiener-Hopf equation (18). Then it follows that

$$N(\tau, x, \sigma) \equiv 0, \quad \sigma = 0, 1, \dots, k \text{ and } x \in \bar{D}, \quad (21)$$

and the optimal matrix kernel function $\tilde{F}(\tau, x, \sigma)$ is unique.

Proof: From (18) we have

$$\begin{aligned} & \sum_{\sigma=0}^k \tilde{F}(\tau, x, \sigma) E[z(\sigma) z'(\alpha)] \\ &= E[u(\tau, x) z'(\alpha)] \\ &= \sum_{\sigma=0}^k (\tilde{F}(\tau, x, \sigma) + N(\tau, x, \sigma)) E[z(\sigma) z'(\alpha)]. \end{aligned}$$

Thus,

$$\sum_{\sigma=0}^k N(\tau, x, \sigma) E[z(\sigma) z'(\alpha)] = 0.$$

Multiplying each side of the above equation by $N'(\tau, x, \alpha)$ and summing from $\alpha = 0$ to $\alpha = k$ yields

$$\sum_{\sigma=0}^k \sum_{\alpha=0}^k N(\tau, x, \sigma) E[z(\sigma) z'(\alpha)] N'(\tau, x, \alpha) = 0.$$

On the other hand, from (8) and (11) we have

$$E[z(\sigma) z'(\alpha)] = H(\sigma) E[u_m(\sigma) u_m'(\alpha)] H'(\alpha) + R(\sigma) \delta_{\sigma\alpha}.$$

Then it follows that

$$\begin{aligned} & \sum_{\sigma=0}^k \sum_{\alpha=0}^k N(\tau, x, \sigma) H(\sigma) E[u_m(\sigma) u_m'(\alpha)] H'(\alpha) N'(\tau, x, \alpha) \\ &+ \sum_{\sigma=0}^k N(\tau, x, \sigma) H(\sigma) R(\sigma) H'(\sigma) N'(\tau, x, \sigma) = 0. \end{aligned}$$

Since both terms on the right side of the above equation are positive-semidefinite because of the positive-definiteness of $R(\sigma)$, a necessary and sufficient condition for the above equation to hold is $N(\tau, x, \sigma) \equiv 0$, $\sigma = 0, 1, \dots, k$ and $x \in \bar{D}$. Thus, the proof of the lemma is complete. Q.E.D.

In order to facilitate the derivation of the optimal estimators, we rewrite (18) in terms of the following corollary.

Corollary 2: The Wiener-Hopf equation (18) is rewritten for the prediction, filtering, and smoothing problems as follows.

1) Prediction ($\tau > k$)

$$\sum_{\sigma=0}^k A(\tau, x, \sigma) E[z(\sigma) z'(\sigma)] = E[u(\tau, x) z'(\alpha)], \quad (22)$$

for $\alpha = 0, 1, \dots, k$ and $x \in \bar{D}$.

2) Filtering ($\tau = k$)

$$\sum_{\sigma=0}^k F(k, x, \sigma) E[z(\sigma) z'(\alpha)] = E[u(k, x) z'(\alpha)] \quad (23)$$

for $\alpha = 0, 1, \dots, k$ and $x \in \bar{D}$.

3) Smoothing ($\tau < k$)

$$\sum_{\sigma=0}^k B(\tau, k, x, \sigma) E[z(\sigma) z'(\alpha)] = E[u(\tau, x) z'(\alpha)] \quad (24)$$

for $\alpha = 0, 1, \dots, k$ and $x \in \bar{D}$.

In what follows, let us denote the estimation error covariance matrix function by $P(\tau, x, y/k)$,

$$P(\tau, x, y/k) = E[\tilde{u}(\tau, x/k) \tilde{u}'(\tau, y/k)]. \quad (25)$$

IV. DERIVATION OF THE OPTIMAL PREDICTOR

In this section, we derive the optimal prediction estimator by using the Wiener-Hopf theory in the previous section.

Theorem 2: The optimal prediction estimator is given by

$$\hat{u}(k+1, x/k) = \mathcal{L}_x \hat{u}(k, x/k) \quad (26)$$

$$\Gamma_\xi \hat{u}(k+1, \xi/k) = S(k+1, \xi), \quad \xi \in \partial D. \quad (27)$$

Proof: From (22) and (1) we have

$$\sum_{\sigma=0}^k A(k+1, x, \sigma) E[z(\sigma) z'(\alpha)] = \mathcal{L}_x E[u(k, x) z'(\alpha)]$$

since $w(k, x)$ is independent of $z(\alpha)$, $\alpha = 0, 1, \dots, k$. From the Wiener-Hopf equation (23) for the optimal filtering

problem we have

$$\sum_{\sigma=0}^k \{A(k+1, x, \sigma) - \mathcal{L}_x F(k, x, \sigma)\} E[z(\sigma)z'(\alpha)] = 0.$$

Defining $\bar{N}(k, x, \sigma)$ by

$$\bar{N}(k+1, x, \sigma) = A(k+1, x, \sigma) - \mathcal{L}_x F(k, x, \sigma),$$

it is clear that $A(k+1, x, \sigma) + \bar{N}(k+1, x, \sigma)$ also satisfies the Wiener-Hopf equation (22). From the uniqueness of $A(k+1, x, \sigma)$ by Lemma 1 it follows that $\bar{N}(k+1, x, \sigma) \equiv 0$, that is,

$$A(k+1, x, \sigma) = \mathcal{L}_x F(k, x, \sigma). \quad (28)$$

Thus, from (13) and (14) we have

$$\hat{u}(k+1, x/k) = \mathcal{L}_x \sum_{\sigma=0}^k F(\sigma, x, \sigma) z(\sigma) = \mathcal{L}_x \hat{u}(k, x/k).$$

Since the forms of Γ_ξ and $S(k+1, \xi)$ are known, the predicted estimate $\hat{u}(k+1, \xi/k)$ also satisfies the same boundary condition as (3), $\Gamma_\xi \hat{u}(k+1, \xi/k) = S(k+1, \xi)$, $\xi \in \partial D$. Thus, the proof of the theorem is complete. Q.E.D.

Theorem 3: The optimal prediction error covariance matrix function $P(k+1, x, y/k)$ is given by

$$P(k+1, x, y/k) = \mathcal{L}_x P(k, x, y/k) \mathcal{L}_y' + \tilde{Q}(k, x, y), \quad (29)$$

$$\Gamma_\xi P(k+1, \xi, y/k) = 0, \quad \xi \in \partial D \quad (30)$$

where

$$\tilde{Q}(k, x, y) = G(k, x) Q(k, x, y) G'(k, y). \quad (31)$$

Proof: From (1), (16), and (26) it follows that

$$\tilde{u}(k+1, x/k) = \mathcal{L}_x \tilde{u}(k, x/k) + G(k, x) w(k, x) \quad (32)$$

and from (3), (16), and (27),

$$\Gamma_\xi \tilde{u}(k+1, \xi/k) = 0, \quad \xi \in \partial D. \quad (33)$$

Then we have from (31) $P(k+1, x, y/k) = E[\tilde{u}(k+1, x/k) \tilde{u}'(k+1, y/k)] = \mathcal{L}_x P(k, x, y/k) \mathcal{L}_y' + \tilde{Q}(k, x, y)$ and from (33), $E[\Gamma_\xi \tilde{u}(k+1, \xi/k) \tilde{u}'(k+1, y/k)] = \Gamma_\xi P(k+1, \xi, y/k) = 0$. Thus, the proof of the theorem is complete. Q.E.D.

V. DERIVATION OF THE OPTIMAL FILTER

Let us derive the optimal filter by using the Wiener-Hopf theorem for the filtering problem. From (23) it follows that

$$\begin{aligned} & F(k+1, x, k+1) E[z(k+1)z'(\alpha)] \\ & + \sum_{\sigma=0}^k F(k+1, x, \sigma) E[z(\sigma)z'(\alpha)] \\ & = E[u(k+1, x)z'(\alpha)] \end{aligned} \quad (34)$$

for $\alpha = 0, 1, \dots, k+1$.

From (1) and the independence of z_0^{k+1} and $w(k+1, x)$, it follows that $E[u(k+1, x)z'(\alpha)] = \mathcal{L}_x E[u(k, x)z'(\alpha)]$. Applying the Wiener-Hopf equation (23) to the right side

of the above equation yields

$$E[u(k+1, x)z'(\alpha)] = \mathcal{L}_x \sum_{\sigma=0}^k F(k, x, \sigma) E[z(\sigma)z'(\alpha)]. \quad (35)$$

Furthermore, from (8) and the whiteness of $v(k+1)$ we have

$$E[z(k+1)z'(\alpha)] = H(k+1) E[u_m(k+1)z'(\alpha)].$$

Let us introduce $\mathcal{L}_*[\cdot]$ and $[\cdot]\mathcal{L}_*' as follows,$

$$\mathcal{L}_*[\cdot] = \begin{bmatrix} \mathcal{L}_x[\cdot] & 0 \\ & \ddots \\ 0 & \mathcal{L}_{x^m}[\cdot] \end{bmatrix} \quad (36)$$

and

$$[\cdot]\mathcal{L}_*' = (\mathcal{L}_*[\cdot])'. \quad (37)$$

Then from (1) and (7) it follows that

$$u_m(k+1) = \mathcal{L}_* u_m(k) + w_m(k) \quad (38)$$

$$\begin{aligned} \tilde{w}_m(k) &= \text{Col}[G(k, x^1)w(k, x^1), \dots, \\ & G(k, x^m)w(k, x^m)]. \end{aligned} \quad (39)$$

Then we have for $\alpha < k+1$, $E[z(k+1)z'(\alpha)] = H(k+1)\mathcal{L}_* E[u_m(k)z'(\alpha)]$. Applying the Wiener-Hopf equation (23) to the right side of the above equation yields

$$\begin{aligned} & E[z(k+1)z'(\alpha)] \\ & = H(k+1) \mathcal{L}_* \sum_{\sigma=0}^k F_m(k, \sigma) E[z(\sigma)z'(\alpha)] \end{aligned} \quad (40)$$

where

$$F_m(k, \sigma) = \begin{bmatrix} F(k, x^1, \sigma) \\ F(k, x^m, \sigma) \end{bmatrix}. \quad (41)$$

Substituting (35) and (40) into (34) yields

$$\sum_{\sigma=0}^k N_\Delta(k, x, \sigma) E[z(\sigma)z'(\alpha)] = 0, \quad \alpha = 0, 1, \dots, k$$

where

$$\begin{aligned} N_\Delta(k, x, \sigma) &= F(k+1, x, k+1) H(k+1) \mathcal{L}_* F_m(k, \sigma) \\ & - \mathcal{L}_x F(k, x, \sigma) + F(k+1, x, \sigma). \end{aligned}$$

Since it is clear that $F(k, x, \sigma) + N_\Delta(k, x, \sigma)$ also satisfies the Wiener-Hopf equation (23), it follows from Lemma 1 that $N_\Delta(k, x, \sigma) \equiv 0$. Thus, we have the following lemma.

Lemma 2: The optimal matrix kernel function $F(k, x, \sigma)$ of the filter is given by

$$\begin{aligned} F(k+1, x, \sigma) &= \mathcal{L}_x F(k, x, \sigma) \\ & - F(k+1, x, k+1) H(k+1) \mathcal{L}_* F_m(k, \sigma), \\ & \sigma = 0, 1, \dots, k. \end{aligned} \quad (42)$$

Theorem 4: The optimal filtering estimate $\hat{u}(k, x/k)$ is

given by

$$\hat{u}(k+1, x/k+1) = \mathcal{L}_x \hat{u}(k, x/k) + F(k+1, x, k+1)v(k+1) \quad (43)$$

$$v(k+1) = z(k+1) - H(k+1)\mathcal{L}_* \hat{u}_m(k/k) \quad (44)$$

$$\hat{u}(0, x/0) = 0 \quad (45)$$

$$\Gamma_\xi \hat{u}(k+1, \xi/k+1) = S(k+1, \xi), \quad \xi \in \partial D \quad (46)$$

where

$$\hat{u}_m(k/k) = \text{Col}[\hat{u}(k, x^1/k), \dots, \hat{u}(k, x^m/k)]. \quad (47)$$

Proof: Using (14) and (42) yields

$$\begin{aligned} \hat{u}(k+1, x/k+1) &= F(k+1, x, k+1)z(k+1) \\ &\quad + \mathcal{L}_x \sum_{\sigma=0}^k F(k, x, \sigma)z(\sigma) \\ &\quad - F(k+1, x, k+1)H(k+1)\mathcal{L}_* \\ &\quad \cdot \sum_{\sigma=0}^k F_m(k, \sigma)z(\sigma). \end{aligned}$$

Again from (14) we have

$$\begin{aligned} \hat{u}(k+1, x/k+1) &= \mathcal{L}_x \hat{u}(k, x/k) \\ &\quad + F(k+1, x, k+1)v(k+1). \end{aligned}$$

Since we have no information at the initial time, it is suitable to assume an initial value of $\hat{u}(k+1, x/k+1)$ as $\hat{u}(0, x/0) = E[u_0(x)] = 0$. Furthermore, since we know the exact forms of Γ_ξ and $S(k+1, \xi)$, the boundary value $\hat{u}(k+1, \xi/k+1)$ also satisfies the same boundary condition as $u(k+1, \xi)$. Thus, we have $\Gamma_\xi \hat{u}(k+1, \xi/k+1) = S(k+1, \xi)$, $\xi \in \partial D$, and the proof of the theorem is complete. Q.E.D.

Note that $v(k+1)$ defined by (44) is rewritten by using the prediction value of (26) as follows.

$$v(k+1) = z(k+1) - H(k+1)\hat{u}_m(k+1/k) \quad (48)$$

or

$$v(k+1) = H(k+1)\tilde{u}_m(k+1/k) + v(k+1) \quad (49)$$

where

$$\tilde{u}_m(k+1/k) = \text{Col}[\hat{u}(k+1, x^1/k), \dots, \hat{u}(k+1, x^m/k)] \quad (50)$$

and

$$\tilde{u}_m(k+1/k) = u_m(k+1) - \hat{u}_m(k+1/k). \quad (51)$$

$v(k+1)$ is termed the innovation process [8], [9].

In order to find the optimal matrix kernel function $F(k+1, x, k+1)$ for the filtering problem, we introduce the following notation,

$$p_m(\tau, x/k) = [p(\tau, x, x^1/k), \dots, p(\tau, x, x^m/k)] \quad (52)$$

and

$$p_{mm}(\tau/k) = \begin{bmatrix} p(\tau, x^1/k) \\ \vdots \\ p(\tau, x^m/k) \end{bmatrix} \cdot \begin{bmatrix} p(\tau, x^1, x^1/k), \dots, p(\tau, x^1, x^m/k) \\ \vdots \\ p(\tau, x^m, x^1/k), \dots, p(\tau, x^m, x^m/k) \end{bmatrix}. \quad (53)$$

Note from the definitions of $p_m(\tau, x/k)$ and $p_{mm}(\tau/k)$ that

$$p_m(\tau, x/k) = E[\tilde{u}(\tau, x/k)\tilde{u}'_m(\tau/k)] \quad (54)$$

and

$$p_{mm}(\tau/k) = E[\tilde{u}_m(\tau/k)\tilde{u}'_m(\tau/k)]. \quad (55)$$

Furthermore, we define the covariance matrix of the innovation process $v(k+1)$ by $\Gamma(k+1/k)$,

$$\Gamma(k+1/k) = E[v(k+1)v'(k+1)]. \quad (56)$$

Then from (49) it follows that

$$\begin{aligned} \Gamma(k+1/k) &= H(k+1)p_{mm}(k+1/k) \\ &\quad \cdot H'(k+1) + R(k+1). \end{aligned} \quad (57)$$

Then the following theorem holds.

Theorem 5: The optimal filtering gain matrix function $F(k+1, x, k+1)$ is given by

$$F(k+1, x, k+1) = p_m(k+1, x/k) \cdot H'(k+1)\Gamma^{-1}(k+1/k) \quad (58)$$

or

$$F(k+1, x, k+1) = p_m(k+1, x/k) \cdot \psi(k+1/k)H'(k+1)R^{-1}(k+1) \quad (59)$$

where

$$\psi(k+1/k) = (I + \tilde{R}(k+1)p_{mm}(k+1/k))^{-1} \quad (60)$$

and

$$\tilde{R}(k+1) = H'(k+1)R^{-1}(k+1)H(k+1). \quad (61)$$

Proof: From the Wiener-Hopf equation (23) it follows that

$$\begin{aligned} F(k+1, x, k+1)E[z(k+1)z'(k+1)] \\ + \sum_{\sigma=0}^k F(k+1, x, \sigma)E[z(\sigma)z'(k+1)] \\ = E[u(k+1, x)z'(k+1)]. \end{aligned}$$

Substituting (42) into the above equation yields

$$F(k+1, x, k+1) \left(E \left[\left\{ z(k+1) - H(k+1) \mathcal{L}_x \cdot \sum_{\sigma=0}^k F_m(k, \sigma) z(\sigma) \right\} z'(k+1) \right] \right) \\ = E \left[\left\{ u(k+1, x) - \mathcal{L}_x \sum_{\sigma=0}^k F(k, x, \sigma) z(\sigma) \right\} z'(k+1) \right].$$

Substituting (14) into the right side of the above equation and using (26) and the orthogonality condition of (20) yields

$$E \left[\left\{ u(k+1, x) - \mathcal{L}_x \sum_{\sigma=0}^k F(k, x, \sigma) z(\sigma) \right\} z'(k+1) \right] \\ = E[\tilde{u}(k+1, x/k) z'(k+1)] \\ = E[\tilde{u}(k+1, x/k) u'_m(k+1)] H'(k+1) \\ = p_m(k+1, x/k) H'(k+1).$$

Using the orthogonality condition of (20) gives

$$E[v(k+1) z'(k+1)] \\ = H(k+1) E[\tilde{u}_m(k+1/k) \cdot u'_m(k+1)] H'(k+1) + R(k+1) \\ = H(k+1) p_{mm}(k+1/k) \cdot H'(k+1) + R(k+1) \\ = \Gamma(k+1/k). \quad (62)$$

Then we have

$$F(k+1, x, k+1) \Gamma(k+1/k) \\ = p_m(k+1, x/k) H'(k+1). \quad (63)$$

Thus (58) is derived. In order to show the equivalence between (58) and (59), we use the following matrix inversion lemma,

$$PH'(HPH' + R)^{-1} = P(I + H'R^{-1}HP)^{-1}H'R^{-1}. \quad (64)$$

From (58) and (64) we have

$$F(k+1, x, k+1) = p_m(k+1, x/k) \cdot \psi(k+1/k) H'(k+1) R^{-1}(k+1).$$

Then (59) is derived, and the proof of the theorem is complete. Q.E.D.

The equation for the optimal filtering error covariance matrix function $p(k+1, x, y/k+1)$ now must be derived.

Theorem 6: The optimal filtering error covariance matrix function $p(k+1, x, y/k+1)$ is given by

$$p(k+1, x, y/k+1) = p(k+1, x, y/k) \\ - p_m(k+1, x/k) H'(k+1) \cdot \Gamma^{-1}(k+1/k) H(k+1) \cdot p'_m(k+1, y/k) \quad (65)$$

or

$$p(k+1, x, y/k+1) = p(k+1, x, y/k) \\ - p_m(k+1, x/k) \psi(k+1/k) \cdot \tilde{R}(k+1) p'_m(k+1, y/k) \quad (66)$$

where

$$p(0, x, y/0) = p_0(x, y) \quad (67)$$

and

$$\Gamma_\xi p(k+1, \xi, y/k+1) = 0, \quad \xi \in \partial D. \quad (68)$$

Proof: From (1) and (43) we have

$$\tilde{u}(k+1, x/k+1) = \tilde{u}(k+1, x/k) \\ - F(k+1, x, k+1) v(k+1) \quad (69)$$

and from (3) and (46)

$$\Gamma_\xi \tilde{u}(k+1, \xi/k+1) = 0, \quad \xi \in \partial D. \quad (70)$$

Using the independence property between $v(k+1)$ and $\tilde{u}(k+1, x/k)$ or $\tilde{u}(k+1, y/k)$ yields from (69),

$$p(k+1, x, y/k+1) \\ = E[\tilde{u}(k+1, x/k+1) \cdot \tilde{u}'(k+1, y/k+1)] \\ = p(k+1, x, y/k) \\ + F(k+1, x, k+1) E[v(k+1) \cdot v'(k+1)] F'(k+1, y, k+1) \\ - F(k+1, x, k+1) H(k+1) \cdot E[\tilde{u}_m(k+1/k) \tilde{u}'(k+1, y/k)] \\ - E[\tilde{u}(k+1, x/k) \tilde{u}'_m(k+1/k)] \cdot H'(k+1) F'(k+1, y, k+1).$$

Using (58) and (63) it follows that

$$p(k+1, x, y/k+1) \\ = p(k+1, x, y/k) \\ - p_m(k+1, x/k) H'(k+1) \cdot F'(k+1, y, k+1) \\ = p(k+1, x, y/k) \\ - p_m(k+1, x/k) H'(k+1) \cdot \Gamma^{-1}(k+1/k) H(k+1) p'_m(k+1, y/k).$$

Thus (65) is derived. The equivalence between (65) and

(66) is easily shown by using (64). Since the initial value $\hat{u}(0, x/0)$ of $\hat{u}(k+1, x/k+1)$ is zero from (45), it is clear that $p(0, x, y/0) = E[\hat{u}(0, y/0)] = p_0(x, y)$. Multiplying each side of (70) by $\hat{u}'(k+1, y/k+1)$ and taking the expectation yields $\Gamma_\xi p(k+1, \xi, y/k+1) = 0, \xi \in \partial D$. Thus, the proof of the theorem is complete. Q.E.D.

Corollary 3: $\hat{u}_m(k+1/k+1)$ and $p_m(k+1, x/k+1)$ satisfy the following relations,

$$\begin{aligned}\hat{u}_m(k+1/k+1) \\ = \hat{u}_m(k+1/k) + F_m(k+1, k+1)v(k+1)\end{aligned}\quad (71)$$

$$\begin{aligned}F_m(k+1, k+1) \\ = p_{mm}(k+1/k)\psi(k+1/k)H'(k+1)R^{-1}(k+1)\end{aligned}\quad (72)$$

or

$$F_m(k+1, k+1) = p_{mm}(k+1/k+1) \cdot H'(k+1)R^{-1}(k+1) \quad (73)$$

$$\begin{aligned}p_{mm}(k+1/k+1) &= p_{mm}(k+1/k) - p_{mm}(k+1/k) \\ &\quad \cdot \psi(k+1/k)\tilde{R}(k+1)p_{mm}(k+1/k)\end{aligned}\quad (74)$$

or

$$p_{mm}(k+1/k+1) = p_{mm}(k+1/k)\psi(k+1/k). \quad (75)$$

Proof: From the definitions (41) and (50) of $F_m(k+1, k+1)$ and $\hat{u}_m(k+1/k)$, it is clear that (71), (72), and (74) hold. From (60) and (74) it follows that

$$\begin{aligned}p_{mm}(k+1/k+1) &= p_{mm}(k+1/k)\psi(k+1/k) \\ &\quad \cdot \{\psi^{-1}(k+1/k) - \tilde{R}(k+1) \\ &\quad \cdot p_{mm}(k+1/k)\} \\ &= p_{mm}(k+1/k)\psi(k+1/k) \\ &\quad \cdot \{I + \tilde{R}(k+1)p_{mm}(k+1/k) \\ &\quad - \tilde{R}(k+1)p_{mm}(k+1/k)\} \\ &= p_{mm}(k+1/k)\psi(k+1/k).\end{aligned}$$

Thus, (75) is derived and (73) is clear from (72) and (75). Q.E.D.

The present result corresponds to that of Santis *et al.* [17] which is an abstract form of the filter.

VI. DERIVATION OF THE EQUATIONS FOR THE OPTIMAL SMOOTHING ESTIMATOR

In this section, we derive the basic equations for the optimal smoothing estimator by using the Wiener-Hopf theory.

Lemma 3: The optimal matrix kernel function $B(\tau, k+1, x, \sigma)$ of the smoothing estimator is given by

$$\begin{aligned}B(\tau, k+1, x, \sigma) &= B(\tau, k, x, \sigma) \\ &\quad - B(\tau, k+1, x, k+1)H(k+1)\mathcal{L}_*F_m(k, \sigma), \\ &\quad \sigma = 0, 1, \dots, k.\end{aligned}\quad (76)$$

Proof: From the Wiener-Hopf equation (24) we have

$$\begin{aligned}\sum_{\sigma=0}^{k+1} B(\tau, k+1, x, \sigma)E[z(\sigma)z'(\alpha)] &= E[u(\tau, x)z'(\alpha)], \\ \alpha &= 0, \dots, k+1\end{aligned}$$

and

$$\begin{aligned}\sum_{\sigma=0}^k B(\tau, k, x, \sigma)E[z(\sigma)z'(\alpha)] &= E[u(\tau, x)z'(\alpha)], \\ \alpha &= 0, \dots, k.\end{aligned}$$

Subtracting the latter equation from the former yields

$$\begin{aligned}B(\tau, k+1, x, k+1)E[z(k+1)z'(\alpha)] \\ + \sum_{\sigma=0}^k (B(\tau, k+1, x, \sigma) \\ - B(\tau, k, x, \sigma))E[z(\sigma)z'(\alpha)] = 0.\end{aligned}$$

From (8) and (23) we have

$$\begin{aligned}E[z(k+1)z'(\alpha)] &= H(k+1)\mathcal{L}_*E[u_m(k)z'(\alpha)] \\ &= H(k+1)\mathcal{L}_*\sum_{\sigma=0}^k F_m(k, \sigma) \\ &\quad \cdot E[z(\sigma)z'(\alpha)].\end{aligned}$$

Then it follows that

$$\sum_{\sigma=0}^k \tilde{N}(\tau, k, x, \sigma)E[z(\sigma)z'(\alpha)] = 0$$

where

$$\begin{aligned}\tilde{N}(\tau, k, x, \sigma) &= B(\tau, k+1, x, \sigma) - B(\tau, k, x, \sigma) \\ &\quad + B(\tau, k+1, x, k+1)H(k+1)\mathcal{L}_*F_m(k, \sigma).\end{aligned}$$

Since it is easily seen that $B(\tau, k, x, \sigma) + \tilde{N}(\tau, k, x, \sigma)$ also satisfies the Wiener-Hopf equation (24), from Lemma 1 we have $\tilde{N}(\tau, k, x, \sigma) \equiv 0$, and the proof of the lemma is complete. Q.E.D.

Theorem 7: The optimal smoothing estimate $\hat{u}(\tau, x/k+1)$ is given by

$$\begin{aligned}\hat{u}(\tau, x/k+1) &= \hat{u}(\tau, x/k) + B(\tau, k+1, x, k+1) \\ &\quad \cdot v(k+1)\end{aligned}\quad (77)$$

$$\Gamma_\xi \hat{u}(\tau, \xi/k+1) = S(\tau, \xi), \xi \in \partial D$$

$$k = \tau, \tau+1, \dots \quad (78)$$

Proof: From (15) it follows that

$$\hat{u}(\tau, x/k+1) = B(\tau, k+1, x, k+1)z(k+1) + \sum_{\sigma=0}^k B(\tau, k+1, x, \sigma)z(\sigma).$$

Substituting (76) into the above equation yields

$$\begin{aligned} \hat{u}(\tau, x/k+1) &= B(\tau, k+1, x, k+1) \\ &\cdot \left(z(k+1) - H(k+1)\mathcal{L}_* \sum_{\sigma=0}^k F_m(k, \sigma)z(\sigma) \right) \\ &+ \sum_{\sigma=0}^k B(\tau, k, x, \sigma)z(\sigma). \end{aligned}$$

Substituting (14) and (15) into the above equation yields

$$\hat{u}(\tau, x/k+1) = \hat{u}(\tau, x/k) + B(\tau, k+1, x, k+1)v(k+1).$$

Since we have no additional information about the boundary value of $u(\tau, x)$, except for $S(\tau, \xi)$ and the exact form Γ_ξ , we have $\Gamma_\xi \hat{u}(\tau, \xi/k+1) = S(\tau, \xi)$, $\xi \in \partial D$, and the proof of the theorem is complete. Q.E.D.

Theorem 8: The optimal smoothing gain matrix function $B(\tau, k+1, x, k+1)$ is given by

$$B(\tau, k+1, x, k+1) = L_m(\tau, x/k)\mathcal{L}'_* \cdot H'(k+1)\Gamma^{-1}(k+1/k) \quad (79)$$

or

$$B(\tau, k+1, x, k+1) = J(\tau, x/k+1) \cdot H'(k+1)R^{-1}(k+1) \quad (80)$$

where

$$J(\tau, x/k+1) = L_m(\tau, x/k)\mathcal{L}'_* \cdot (I + \tilde{R}(k+1)p_{mm}(k+1/k))^{-1} \quad (81)$$

$$L_m(\tau, x/k) = [L(\tau, x, x^1/k), \dots, L(\tau, x, x^m/k)] \quad (82)$$

and

$$L(\tau, x, y/k) = E[\tilde{u}(\tau, x/k)\tilde{u}'(k, y/k)]. \quad (83)$$

Proof: From the Wiener-Hopf equation (24) it follows that

$$\begin{aligned} B(\tau, k+1, x, k+1)E[z(k+1)z'(k+1)] \\ + \sum_{\sigma=0}^k B(\tau, k+1, x, \sigma)E[z(\sigma)z'(k+1)] \\ = E[u(\tau, x)z'(k+1)]. \end{aligned}$$

Substituting (76) into the above equation yields

$$\begin{aligned} B(\tau, k+1, x, k+1)E[v(k+1)z'(k+1)] \\ = E[\tilde{u}(\tau, x/k)z'(k+1)]. \end{aligned}$$

On the other hand, from (48) and (49) we have

$$\begin{aligned} E[v(k+1)z'(k+1)] &= E[v(k+1)(v(k+1) \\ &\quad + H(k+1)\hat{u}_m(k+1/k))'] \\ &= E[v(k+1)v'(k+1)] \\ &= \Gamma(k+1/k). \end{aligned} \quad (84)$$

From (8) and the independence of $v(k+1)$ and $\tilde{u}(\tau, x/k)$, we have

$$\begin{aligned} E[\tilde{u}(\tau, x/k)z'(k+1)] \\ = E[\tilde{u}(\tau, x/k)\tilde{u}'_m(k+1/k)]H'(k+1). \end{aligned}$$

But from (26) and (38) it follows that

$$\tilde{u}_m(k+1/k) = \mathcal{L}_*\tilde{u}_m(k/k) + \tilde{w}_m(k). \quad (85)$$

Then we have

$$\begin{aligned} B(\tau, k+1, x, k+1)\Gamma(k+1/k) \\ = L_m(\tau, x/k)\mathcal{L}'_*H'(k+1). \end{aligned}$$

The equivalence between (79) and (80) is easily seen by using the matrix inversion lemma (64). Thus the proof of the theorem is complete. Q.E.D.

Let us now derive the equation for $L(\tau, x, y/k+1)$. Using the orthogonality condition (20) yields

$$L(\tau, x, y/k+1) = E[u(\tau, x)\tilde{u}'(k+1, y/k+1)].$$

Substituting (69) into the above equation yields

$$\begin{aligned} L(\tau, x, y/k+1) &= L(\tau, x, y/k)\mathcal{L}'_* \\ &\quad - L_m(\tau, x/k)\mathcal{L}'_*H'(k+1)F'(k+1, y, k+1). \end{aligned} \quad (86)$$

From (3) and (78) it follows that $\Gamma_\xi \tilde{u}(\tau, \xi/k+1) = 0$, $\xi \in \partial D$. Multiplying each side by $\tilde{u}'(k+1, y/k+1)$ and taking the expectation yields

$$\Gamma_\xi L(\tau, \xi, y/k+1) = 0, \quad \xi \in \partial D. \quad (87)$$

Then the following theorem holds.

Theorem 9: $J(\tau, x/k+1)$ in (80) is given by

$$J(\tau, x/k+1) = J(\tau, x/k)\mathcal{L}'_*\psi(k+1/k) \quad (88)$$

$$J(\tau, x/\tau) = p_m(\tau, x/\tau) \quad (89)$$

$$\Gamma_\xi J(\tau, \xi/k+1) = 0, \quad \xi \in \partial D. \quad (90)$$

Proof: From (86) and (59) it follows that

$$\begin{aligned} L_m(\tau, x/k+1) &= L_m(\tau, x)\mathcal{L}'_* \\ &\quad \cdot \{I - \tilde{R}(k+1)\psi'(k+1/k)p_{mm}(k+1/k)\}. \end{aligned}$$

But we have

$$\begin{aligned} I - \tilde{R}(I + P\tilde{R})^{-1}P &= \tilde{R}(I + P\tilde{R})^{-1} \\ &\quad \cdot ((I + P\tilde{R})\tilde{R}^{-1} - P) \\ &= ((I + P\tilde{R})R^{-1})^{-1}\tilde{R}^{-1} \\ &= (\tilde{R}(\tilde{R}^{-1} + P))^{-1} \\ &= (I + \tilde{R}P)^{-1}. \end{aligned}$$

Thus,

$$L_m(\tau, x/k+1) = L_m(\tau, x) \mathcal{L}'_* \cdot (I + \tilde{R}(k+1)p_{mm}(k+1/k))^{-1}.$$

Therefore, from (81) it follows that

$$J(\tau, x/k+1) = L_m(\tau, x/k+1) \quad (91)$$

and from (81) we have

$$J(\tau, x/k+1) = J(\tau, x/k) \mathcal{L}'_* \cdot (I + \tilde{R}(k+1)p_{mm}(k+1/k))^{-1}.$$

Then it follows that

$$J(\tau, x/\tau) = L_m(\tau, x/\tau) = p_m(\tau, x/\tau).$$

Since (90) is clear from (87) and (91), the proof of the theorem is complete. Q.E.D.

Let us now derive the equation for the optimal smoothing error covariance matrix function $p(\tau, x, y/k)$ defined by

$$p(\tau, x, y/k) = E[\tilde{u}(\tau, x/k)\tilde{u}(\tau, y/k)]. \quad (92)$$

From (77) and (78) it follows that

$$\tilde{u}(\tau, x/k+1) = \tilde{u}(\tau, x/k) - B(\tau, k+1, x, k+1)v(k+1) \quad (93)$$

$$\Gamma_\xi \tilde{u}(\tau, \xi/k+1) = 0, \quad \xi \in \partial D. \quad (94)$$

Then the following theorem holds.

Theorem 10: The optimal smoothing error covariance matrix function $p(\tau, x, y/k+1)$ is given by

$$\begin{aligned} p(\tau, x, y/k+1) &= p(\tau, x, y/k) \\ &\quad - L_m(\tau, x/k) \mathcal{L}'_* H'(k+1) \\ &\quad \cdot \Gamma^{-1}(k+1/k) H(k+1) \mathcal{L}'_* L'_m(\tau, y/k) \end{aligned} \quad (95)$$

or

$$\begin{aligned} p(\tau, x, y/k+1) &= p(\tau, x, y/k) \\ &\quad - J(\tau, x/k+1) \\ &\quad \cdot \psi^{-1}(k+1/k) \tilde{R}(k+1) J'(\tau, y/k+1) \end{aligned} \quad (96)$$

$$\Gamma_\xi p(\tau, \xi, y/k+1) = 0, \quad \xi \in \partial D. \quad (97)$$

Proof: From (93) it follows that

$$\begin{aligned} p(\tau, x, y/k+1) &= p(\tau, x, y/k) \\ &\quad + B(\tau, k+1, x, k+1) E[v(k+1) \\ &\quad \cdot v'(k+1)] B'(\tau, k+1, y, k+1) \\ &\quad - B(\tau, k+1, x, k+1) E[v(k+1) \\ &\quad \cdot \tilde{u}'(\tau, y/k)] - E[\tilde{u}(\tau, x/k) v'(k+1)] \\ &\quad \cdot B'(\tau, k+1, y, k+1). \end{aligned}$$

$$\begin{aligned} E[\tilde{u}(\tau, x/k) v'(k+1)] &= E[\tilde{u}(\tau, x/k) \tilde{u}'_m(k/k)] \\ &\quad \cdot \mathcal{L}'_* H'(k+1) \\ &= L_m(\tau, x/k) \mathcal{L}'_* H'(k+1) \end{aligned}$$

and

$$E[v(k+1) \tilde{u}'(\tau, y/k)] = H(k+1) \mathcal{L}'_* L'_m(\tau, y/k).$$

Thus, we have

$$\begin{aligned} p(\tau, x, y/k+1) &= p(\tau, x, y/k) \\ &\quad + B(\tau, k+1, x, k+1) \\ &\quad \cdot \Gamma(k+1/k) B'(\tau, k+1, y, k+1) \\ &\quad - B(\tau, k+1, x, k+1) H(k+1) \\ &\quad \cdot \mathcal{L}'_* L'_m(\tau, y/k) - L_m(\tau, x/k) \\ &\quad \cdot \mathcal{L}'_* H'(k+1) B'(\tau, k+1, y, k+1). \end{aligned}$$

Substituting (79) into the above equation yields

$$\begin{aligned} p(\tau, x, y/k+1) &= p(\tau, x, y/k) - L_m(\tau, x/k) \mathcal{L}'_* \\ &\quad \cdot H'(k+1) \Gamma^{-1}(k+1/k) H(k+1) \mathcal{L}'_* L'_m(\tau, y/k). \end{aligned}$$

In order to derive (96), note that from (81),

$$L_m(\tau, x/k) \mathcal{L}'_* = J(\tau, x/k+1) \psi^{-1}(k+1/k)$$

and from the matrix inversion lemma (64),

$$H'(HPH' + R)^{-1}H = (I + H'R^{-1}HP)^{-1}H'R^{-1}H.$$

Then we have

$$\begin{aligned} H'(k+1) \Gamma^{-1}(k+1/k) H(k+1) \\ = \psi(k+1/k) \tilde{R}(k+1) \end{aligned}$$

and

$$\begin{aligned} p(\tau, x, y/k+1) &= p(\tau, x, y/k) - J(\tau, x/k+1) \\ &\quad \cdot \psi^{-1}(k+1/k) \tilde{R}(k+1) J'(\tau, y/k+1). \end{aligned}$$

Multiplying each side of (94) by $\tilde{u}(\tau, y/k+1)$ and taking the expectation yields $\Gamma_\xi p(\tau, \xi, y/k+1) = 0, \xi \in \partial D$. Thus, the proof of the theorem is complete. Q.E.D.

Corollary 4: $J(\tau, x/k)$ satisfies the following relations.

$$J(\tau, x/k+1) = A(\tau, x) J_m(\tau+1/k+1) \quad (98)$$

and

$$J(\tau+1, x/k) = D(\tau, x) J_m(\tau/k) \quad (99)$$

where

$$J_m(\tau/k) = \begin{bmatrix} J(\tau, x^1/k) \\ \vdots \\ J(\tau, x^m/k) \end{bmatrix} \quad (100)$$

$$A(\tau, x) = p_m(\tau, x/\tau) \mathcal{L}'_* p_{mm}^{-1}(\tau+1/\tau) \quad (101)$$

$$D(\tau, x) = p_m(\tau+1, x/\tau) (p_{mm}(\tau/\tau) \mathcal{L}'_*)^{-1}. \quad (102)$$

Proof: Letting $\Phi(k+1)$ be given by $\Phi(k+1) = \mathcal{L}'_*(I + \tilde{R}(k+1)p_{mm}(k+1/k))^{-1}$, from (88) and (89) it

follows that $J(\tau, x/k+1) = p_m(\tau, x/\tau)\Phi(\tau+1)\Phi(\tau+2)\cdots\Phi(k+1)$ and $J_m(\tau+1/k+1) = p_{mm}(\tau+1/\tau+1)\Phi(\tau+2)\cdots\Phi(k+1)$. From the above equations and (75) we have

$$\begin{aligned} J(\tau, x/k+1) &= p_m(\tau, x/\tau)\Phi(\tau+1) \\ &\quad \cdot p_{mm}^{-1}(\tau+1/\tau+1)J_m(\tau+1/k+1) \\ &= p_m(\tau, x/\tau)\mathcal{L}'_*\psi(\tau+1/\tau)\psi^{-1}(\tau+1/\tau) \\ &\quad \cdot p_{mm}^{-1}(\tau+1/\tau)J_m(\tau+1/k+1) \\ &= A(\tau, x)J_m(\tau+1/k+1). \end{aligned}$$

From (88) and (89) it follows that $J(\tau+1, x/k) = p_m(\tau+1, x/\tau+1)\Phi(\tau+2)\cdots\Phi(k)$ and $J_m(\tau/k) = p_{mm}(\tau/\tau)\Phi(\tau+1)\Phi(\tau+2)\cdots\Phi(k)$. Thus, we have from the above equations

$$\begin{aligned} J(\tau+1, x/k) &= p_m(\tau+1, x/\tau+1)(I + \tilde{R}(\tau+1) \\ &\quad \cdot p_{mm}(\tau+1/\tau))(\mathcal{L}'_*\psi(\tau+1/\tau))^{-1}J_m(\tau/k) \\ &= p_m(\tau+1, x/\tau)(p_{mm}(\tau/\tau)\mathcal{L}'_*\psi(\tau+1/\tau))^{-1}J_m(\tau/k) \end{aligned}$$

where the following equality derived from (66) has been used,

$$p_m(\tau+1, x/\tau+1) = p_m(\tau+1, x/\tau) \cdot (I + \tilde{R}(\tau+1)p_{mm}(\tau+1/\tau))^{-1}. \quad (103)$$

Thus, the proof of the corollary is complete. Q.E.D.

Theorem 11: The optimal smoothing estimator is given by

$$\hat{u}(\tau, x/k) = \hat{u}(\tau, x/\tau) + \sum_{l=\tau+1}^k J(\tau, x/l)\bar{v}(l) \quad (104)$$

$$\Gamma_\xi \hat{u}(\tau, \xi/k) = S(\tau, \xi), \quad \xi \in \partial D \quad (105)$$

where

$$\bar{v}(l) = H'(l)R^{-1}(l)v(l). \quad (106)$$

Furthermore, the optimal smoothing error covariance matrix function $p(\tau, x, y/k)$ is given by

$$\begin{aligned} p(\tau, x, y/k) &= p(\tau, x, y/\tau) - \sum_{l=\tau+1}^k J(\tau, x/l) \\ &\quad \cdot \psi^{-1}(l/l-1)\tilde{R}(l)J'(\tau, y/l) \quad (107) \end{aligned}$$

$$\Gamma_\xi p(\tau, \xi, y/k) = 0, \quad \xi \in \partial D. \quad (108)$$

Proof: From (77) and (80), (104) can be directly obtained and from (96), (107) is clear. Thus, the proof of the theorem is complete. Q.E.D.

VII. SUMMARY OF THE OPTIMAL SMOOTHING ESTIMATORS

A. Fixed-Point Smoother

(τ = fixed, $k = \tau+1, \tau+2, \dots$)

Theorem 12: The optimal fixed-point smoothing estimator is given by

$$\hat{u}(\tau, x/k+1) = \hat{u}(\tau, x/k) + J(\tau, x/k+1)\bar{v}(k+1) \quad (109)$$

$$J(\tau, x/k+1) = J(\tau, x/k)\mathcal{L}'_*\psi(k+1/k) \quad (110)$$

$$\psi(k+1/k) = (I + \tilde{R}(k+1)p_{mm}(k+1))^{-1} \quad (111)$$

$$J(\tau, x/\tau) = p_m(\tau, x/\tau) \quad (112)$$

$$\Gamma_\xi \hat{u}(\tau, \xi/k+1) = S(\tau, \xi), \quad \xi \in \partial D \quad (113)$$

$$\Gamma_\xi J(\xi, x/k+1) = 0, \quad \xi \in \partial D. \quad (114)$$

Furthermore, the optimal fixed-point smoothing error covariance matrix function $p(\tau, x, y/k+1)$ is given by

$$\begin{aligned} p(\tau, x, y/k+1) &= p(\tau, x, y/k) - J(\tau, x/k+1) \\ &\quad \cdot \psi^{-1}(k+1/k)\tilde{R}(k+1)J'(\tau, y/k+1) \quad (115) \end{aligned}$$

$$\Gamma_\xi p(\tau, \xi, y/k+1) = 0, \quad \xi \in \partial D. \quad (116)$$

B. Fixed-Interval Smoothing Estimator

(k = fixed, $\tau = k-1, k-2, \dots$)

From Theorem 11 it follows that

$$\begin{aligned} \hat{u}(\tau+1, x/k) &= \hat{u}(\tau+1, x/\tau+1) \\ &\quad + \sum_{l=\tau+2}^k J(\tau+1, x/l)\bar{v}(l) \quad (117) \end{aligned}$$

and

$$\begin{aligned} p(\tau+1, x, y/k) &= p(\tau+1, x, y/\tau+1) \\ &\quad - \sum_{l=\tau+2}^k J(\tau+1, x/l)\psi^{-1}(l/l-1) \\ &\quad \cdot \tilde{R}(l)J'(\tau+1, y/l). \quad (118) \end{aligned}$$

Then the following theorem holds.

Theorem 13: The optimal fixed-interval smoothing estimator is given by

$$\begin{aligned} \hat{u}(\tau+1, x/k) &= \hat{u}(\tau+1, x/\tau+1) \\ &\quad + A(\tau+1, x)[\hat{u}_m(\tau+2/k) \\ &\quad - \hat{u}_m(\tau+2/\tau+1)] \quad (119) \end{aligned}$$

$$\Gamma_\xi \hat{u}(\tau+1, \xi/k) = S(\tau+1, \xi), \quad \xi \in \partial D. \quad (120)$$

Furthermore, the optimal fixed-interval smoothing error covariance matrix function is given by

$$\begin{aligned} p(\tau+1, x, y/k) &= p(\tau+1, x, y/\tau+1) \\ &\quad - A(\tau+1, x)(p_{mm}(\tau+1/k) \\ &\quad - p_{mm}(\tau+1/\tau))A'(\tau+1, y) \quad (121) \end{aligned}$$

$$\Gamma_\xi p(\tau+1, \xi, y/k) = 0, \quad \xi \in \partial D. \quad (122)$$

Proof: From (98) and (117) we have

$$\begin{aligned} \hat{u}(\tau+1, x/k) &= \hat{u}(\tau+1, x/\tau+1) \\ &\quad + A(\tau+1, x) \sum_{l=\tau+2}^k J_m(\tau+1/l)\bar{v}(l). \end{aligned}$$

But from Theorem 11,

$$\hat{u}(\tau + 2, x/k) = \hat{u}(\tau + 2, x/\tau + 2) + \sum_{l=\tau+3}^k J(\tau + 2, x/l) \bar{p}(l)$$

and from (43) and (59),

$$\begin{aligned} \hat{u}(\tau + 2, x/\tau + 2) &= \hat{u}(\tau + 2, x/\tau + 1) \\ &\quad + F(\tau + 2, x, \tau + 2) v(\tau + 2) \\ &= \hat{u}(\tau + 2, x/\tau + 1) \\ &\quad + J(\tau + 2, x/\tau + 2) \bar{p}(\tau + 2). \end{aligned}$$

Thus, we have

$$\begin{aligned} \hat{u}_m(\tau + 2/k) - \hat{u}_m(\tau + 2/\tau + 1) \\ = \sum_{l=\tau+2}^k J_m(\tau + 2/l) \bar{p}(l). \end{aligned}$$

Then we have

$$\begin{aligned} \hat{u}(\tau + 1, x/k) &= \hat{u}(\tau + 1, x/\tau + 1) + A(\tau + 1, x) \\ &\quad \cdot [\hat{u}_m(\tau + 2/k) - \hat{u}_m(\tau + 2/\tau + 1)]. \end{aligned}$$

From (98) and (118),

$$\begin{aligned} p(\tau + 1, x, y/k) &= p(\tau + 1, x, y/\tau + 1) \\ &\quad - A(\tau + 1, x) \sum_{l=\tau+2}^k J_m(\tau + 2/l) \\ &\quad \cdot \psi^{-1}(l/l - 1) \tilde{R}(l) J'_m(\tau + 2/l) \\ &\quad \cdot A'(\tau + 1, y). \end{aligned}$$

From Theorem 11,

$$\begin{aligned} p(\tau + 2, x, y/k) &= p(\tau + 2, x, y/\tau + 2) \\ &\quad - \sum_{l=\tau+3}^k J(\tau + 2, x/l) \psi^{-1}(l/l - 1) \tilde{R}(l) J'(\tau + 2, y/l) \end{aligned}$$

and from (66),

$$\begin{aligned} p(\tau + 2, x, y/\tau + 2) \\ = p(\tau + 2, x, y/\tau + 1) \\ - p_m(\tau + 2, x/\tau + 1) \psi(\tau + 2/\tau + 1) \\ \cdot \tilde{R}(\tau + 2) p'_m(\tau + 2, y/\tau + 1). \end{aligned}$$

Taking into consideration that, from (103), $J(\tau + 2, x/\tau + 2) = p_m(\tau + 2, x/\tau + 1) \psi(\tau + 2/\tau + 1)$ and

$$\begin{aligned} p_{mm}(\tau + 2/k) - p_{mm}(\tau + 2/\tau + 1) \\ = - \sum_{l=\tau+2}^k J_m(\tau + 2/l) \psi^{-1}(l/l - 1) \tilde{R}(l) J'_m(\tau + 2/l), \end{aligned}$$

we have $p(\tau + 1, x, y/k) = p(\tau + 1, x, y/\tau + 1) - A(\tau + 1, x)[p_{mm}(\tau + 2/k) - p_{mm}(\tau + 2/\tau + 1)]A'(\tau + 1, y)$. Since the boundary conditions (120) and (122) are clear from (105) and (108), respectively, the proof of the theorem is complete. Q.E.D.

C. Fixed-Lag Smoothing Estimator

($\tau = k + 1, k = k + 1 + \Delta, \Delta = \text{fixed}$)

From Theorem 11 we have

$$\begin{aligned} \hat{u}(k + 1, x/k + 1 + \Delta) \\ = \hat{u}(k + 1, x/k + 1) \\ + \sum_{l=k+2}^{k+1+\Delta} J(k + 1, x/l) \bar{p}(l) \end{aligned} \quad (123)$$

$$\begin{aligned} \cdot p(k + 1, x, y/k + 1 + \Delta) \\ = p(k + 1, x, y/k + 1) \\ - \sum_{l=k+2}^{k+1+\Delta} J(k + 1, x/l) \\ \cdot \psi^{-1}(l/l - 1) \tilde{R}(l) J'(k + 1, y/l). \end{aligned} \quad (124)$$

Then the following theorem holds.

Theorem 14: The optimal fixed-lag smoothing estimator is given by

$$\begin{aligned} \hat{u}(k + 1, x/k + 1 + \Delta) \\ = \mathcal{L}_x \hat{u}(k, x/k + \Delta) + C(x, k + 1, \Delta) \\ \cdot F_m(k + 1 + \Delta/k + 1 + \Delta) v(k + 1 + \Delta) \\ + \tilde{Q}_m(k, x) (p_{mm}(k/k) \mathcal{L}'_*)^{-1} \\ \cdot (\hat{u}_m(k/k + \Delta) - \hat{u}_m(k/k)) \end{aligned} \quad (125)$$

$$\Gamma_\xi \hat{u}(k + 1, \xi/k + 1 + \Delta) = S(k + 1, \xi), \quad \xi \in \partial D \quad (126)$$

where

$$\begin{aligned} C(x, k + 1, \Delta) &= A(k + 1, x) A_m(k + 1), \\ &\quad \dots, A_m(k + \Delta) \end{aligned} \quad (127)$$

and

$$A_m(k) = \begin{bmatrix} A(k, x^1) \\ \vdots \\ A(k, x^m) \end{bmatrix}.$$

Furthermore, the optimal fixed-lag smoothing error covariance matrix function $p(k + 1, x, y/k + 1 + \Delta)$ is given by

$$\begin{aligned} p(k + 1, x, y/k + 1 + \Delta) \\ = p(k + 1, x, y/k) - C(x, k + 1, \Delta) \\ \cdot F_m(k + 1 + \Delta/k + 1 + \Delta) H(k + 1 + \Delta) \\ \cdot p_{mm}(k + 1 + \Delta/k + \Delta) C'(y, k + 1, \Delta) - D(k, x) \\ \cdot [p_{mm}(k/k) - p_{mm}(k/k + \Delta)] D'(k, y) \end{aligned} \quad (128)$$

$$\Gamma_\xi p(k + 1, \xi, y/k + 1 + \Delta) = 0, \quad \xi \in \partial D. \quad (129)$$

Proof: From (43) and (59) we have

$$\begin{aligned} \hat{u}(k + 1, x/k + 1) &= \mathcal{L}_x \hat{u}(k, x/k) \\ &\quad + J(k + 1, x/k + 1) \bar{p}(k + 1). \end{aligned}$$

From (123) and the above equation it follows that

$$\begin{aligned}\hat{u}(k+1, x/k+1+\Delta) &= \mathcal{L}_x \hat{u}(k, x/k) \\ &+ \sum_{l=k+1}^{k+\Delta} J(k+1, x/l) \bar{v}(l) \\ &+ J(k+1, x/k+1+\Delta) \bar{v}(k+1+\Delta).\end{aligned}$$

From (88) it follows that

$$\begin{aligned}J(k+1, x/k+1+\Delta) &= p_m(k+1, x/k+1) \\ &\cdot \mathcal{L}'_* \psi(k+2/k+1) \\ &\cdot \mathcal{L}'_* \psi(k+3/k+2) \cdots \\ &\cdot \mathcal{L}'_* \psi(k+2+\Delta/k+1+\Delta).\end{aligned}$$

Substituting (75) into the right side yields

$$\begin{aligned}J(k+1, x/k+1+\Delta) &= p_m(k+1, x/k+1) \mathcal{L}'_* p_{mm}^{-1}(k+2/k+1) \\ &\cdot p_{mm}(k+2/k+2) \mathcal{L}'_* \psi(k+3/k+2) \cdots \\ &\cdot \mathcal{L}'_* \psi(k+2+\Delta/k+1+\Delta).\end{aligned}$$

Repeating the same procedure and using (101) yields

$$\begin{aligned}J(k+1, x/k+1+\Delta) &= A(k+1, x) A_m(k+2) \cdots A_m(k+\Delta) \\ &\cdot p_{mm}(k+1+\Delta/k+1+\Delta).\end{aligned}$$

Thus we have

$$\begin{aligned}J(k+1, x/k+1+\Delta) \bar{v}(k+1+\Delta) &= C(x, k+1, \Delta) \\ &\cdot p_{mm}(k+1+\Delta/k+1+\Delta).\end{aligned}\quad (130)$$

From (99) it follows that

$$\begin{aligned}\sum_{l=k+1}^{k+\Delta} J(k+1, x/l) \bar{v}(l) &= \sum_{l=k+1}^{k+\Delta} p_m(k+1, x/k) \\ &\cdot (p_{mm}(k/k) \mathcal{L}'_*)^{-1} J_m(k/l) \bar{v}(l).\end{aligned}$$

But from (29) we have

$$p_m(k+1, x/k) = \mathcal{L}_x p_m(k, x/k) \mathcal{L}'_* + Q_m(k, x).$$

From (98) and (99) we have

$$\begin{aligned}J(k, x/l) &= A(k, x) J_m(k+1/l) = p_m(\tau, x/\tau) \\ &\cdot \mathcal{L}'_* (p_{mm}(k/k) \mathcal{L}'_*)^{-1} J_m(k/l).\end{aligned}$$

Then it follows that

$$\begin{aligned}\sum_{l=k+1}^{k+\Delta} J(k+1, x/l) \bar{v}(l) &= \mathcal{L}_x \sum_{l=k+1}^{k+\Delta} J(k, x/l) \bar{v}(l) \\ &+ \tilde{Q}_m(k, x) \sum_{l=k+1}^{k+\Delta} (p_{mm}(k/k) \mathcal{L}'_*)^{-1} J_m(k/l) \bar{v}(l)\end{aligned}$$

and

$$\begin{aligned}\hat{u}(k+1, x/k+1+\Delta) &= \mathcal{L}_x \hat{u}(k, x/k+\Delta) + C(x, k+1, \Delta) \\ &\cdot F_m(k+1+\Delta/k+1+\Delta) \bar{v}(k+1+\Delta) \\ &+ \tilde{Q}_m(k, x) (p_{mm}(k/k) \mathcal{L}'_*)^{-1} \sum_{l=l+1}^{k+\Delta} J_m(k/l) \bar{v}(l).\end{aligned}$$

But from Theorem 11 we have

$$\hat{u}_m(k/k+\Delta) - \hat{u}_m(k/k) = \sum_{l=k+1}^{k+\Delta} J_m(k/l) \bar{v}(l).$$

Thus we have (125). From (65) and (124) it follows that

$$p(k+1, x, y/k+1+\Delta) = p(k+1, x, y/k) - J_1 - J_2$$

where

$$\begin{aligned}J_1 &= \sum_{l=k+1}^{k+\Delta} J(k+1, x/l) \psi^{-1}(l/l-1) \tilde{R}(l) \\ &\cdot J'(k+1, y/l) \\ J_2 &= J(k+1, x/k+1+\Delta) \psi^{-1}(k+1+\Delta/k+\Delta) \\ &\cdot \tilde{R}(k+1+\Delta) J'(k+1, y/k+1+\Delta).\end{aligned}$$

From (75) and (130) we have

$$\begin{aligned}J_2 &= C(x, k+1, \Delta) p_{mm}(k+1+\Delta/k+1+\Delta) \\ &\cdot \tilde{R}(k+1+\Delta) p_{mm}(k+1+\Delta/k+\Delta) \\ &\cdot C'(y, k+1, \Delta) \\ &= C(x, k+1, \Delta) F_m(k+1+\Delta/k+1+\Delta) \\ &\cdot H(k+1+\Delta) p_{mm}(k+1+\Delta/k+\Delta) \\ &\cdot C'(y, k+1, \Delta).\end{aligned}$$

Substituting (99) into J_1 yields

$$\begin{aligned}J_1 &= D(k, x) \sum_{l=k+1}^{k+\Delta} J_m(k/l) \psi^{-1}(l/l-1) \\ &\cdot \tilde{R}(l) J'_m(k/l) D'(k, y).\end{aligned}$$

But from Theorem 11 we have

$$\begin{aligned}p(k, x, y/k+\Delta) - p(k, x, y/k) &= - \sum_{l=k+1}^{k+\Delta} J(k, x/l) \psi^{-1}(l/l-1) \tilde{R}(l) J'(k, y/l)\end{aligned}$$

and

$$\begin{aligned}p_{mm}(k/k+\Delta) - p_{mm}(k/k) &= - \sum_{l=k+1}^{k+\Delta} J_m(k/l) \psi^{-1}(l/l-1) \tilde{R}(l) J'_m(k/l).\end{aligned}$$

Then we have

$$\begin{aligned}p(k+1, x, y/k+1+\Delta) &= p(k+1, x, y/k) - C(x, k+1, \Delta) \\ &\cdot F_m(k+1+\Delta/k+1+\Delta) H(k+1+\Delta) \\ &\cdot p_{mm}(k+1+\Delta/k+\Delta) C'(y, k+1, \Delta) - D(k, x) \\ &\cdot [p_{mm}(k/k) - p_{mm}(k/k+\Delta)] D'(k, y).\end{aligned}$$

Since the boundary conditions (126) and (129) are clear from (105) and (108), respectively, the proof of the theorem is complete. Q.E.D.

Kelly and Anderson [18] proved that the fixed-lag smoothing algorithm of Theorem 14 may be unstable, but Chirarattananon and Anderson [19] derived a stable version of the algorithm. It is possible to derive a comparable version here, although stability problems should not arise

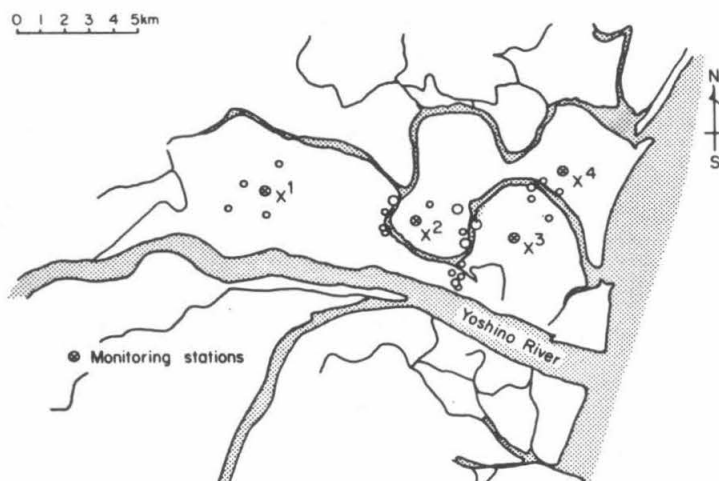


Fig. 1. Map of Tokushima Prefecture, Japan. The four air pollution monitoring stations shown are located as follows. x^1 Aizumi, x^2 Kitajima, x^3 Kawauchi, x^4 Matsushige. Sources of sulfur dioxide have been lumped according to the three sources sizes indicated by the open circles. \bigcirc 30–50 m^3/h \bigcirc 10–30 m^3/h \bigcirc < 10 m^3/h .

in our use of the algorithm of Theorem 14 as long as it is used over a finite time interval.

VIII. APPLICATION TO ESTIMATION OF AIR POLLUTION

Distributed parameter estimation theory has recently been applied to simulated air pollution data to demonstrate the capability of estimating atmospheric concentration levels from routine monitoring data [10], [11]. A problem identified in these early studies was how to specify the statistical properties of the assumed system and observation noise. In this section we expand upon the prior studies in two respects. First, we consider actual monitoring data for sulfur dioxide (SO_2), in particular those measured each hour during the period December 1–31, 1975 at four locations in Tokushima Prefecture, Japan (see Fig. 1). Second, we apply the method of Sage and Husa [12] to estimate the unknown noise covariances in the system equation and measurements.

Hourly sulfur dioxide data are available at the four locations shown in Fig. 1 for the period December 1–31, 1975. The data for day k at location i may be denoted by $e_k(x^i, t)$. It is useful to average the data for December 1–30 to produce

$$\langle e(x^i, t) \rangle = \frac{1}{30} \sum_{k=1}^{30} e_k(x^i, t) \quad (131)$$

where we will consider December 31 as a day to test the algorithms.

If it can be assumed that the wind flows are such that there are no north-south variations of concentration and that vertical mixing is rapid enough to eliminate variations of concentration with altitude, then the region can be considered to be one-dimensional along the east-west coordinate. The SO_2 concentration at any particular time can be assumed to be described by the atmospheric diffusion equation [13],

$$\frac{\partial c}{\partial t} + \zeta \frac{\partial c}{\partial x} = a \frac{\partial^2 c}{\partial x^2} + S(x, t) \quad (132)$$

where ζ is the wind velocity, a is a diffusion coefficient, and S is the rate of emission of SO_2 as a function of location and time.

Equation (132) holds at any instant of time, but we desire an equation governing the monthly mean concentration $\langle c \rangle$. Although no such equation exists, we can formally average (132) over the 30 realizations (days) to produce

$$\frac{\partial \langle c \rangle}{\partial t} + \left\langle \zeta \frac{\partial c}{\partial x} \right\rangle = \left\langle a \frac{\partial^2 c}{\partial x^2} \right\rangle + S. \quad (133)$$

One object will be to estimate the diffusion parameter a . This parameter will in general vary with location and time of day, although for simplicity we seek a constant value for the month. Thus, the first term on the right side of (133) becomes $a \frac{\partial^2 \langle u \rangle}{\partial x^2}$. We can form the residuals, $u = c - \langle c \rangle$ and $z = e - \langle e \rangle$. By subtracting (133) from (132) we obtain

$$\frac{\partial u}{\partial t} + \zeta \frac{\partial c}{\partial x} - \left\langle \zeta \frac{\partial c}{\partial x} \right\rangle = a \frac{\partial^2 u}{\partial x^2}. \quad (134)$$

Since wind data are not available with which to evaluate the second and third terms on the left side of (134) let us rewrite (134) as

$$\frac{\partial u}{\partial t} = a \frac{\partial^2 u}{\partial x^2} + w(x, t) \quad (135)$$

where $w(x, t)$ includes those unknown features associated with the velocity terms.

The boundary conditions on (132) are

$$\frac{\partial c}{\partial x} = 0, \quad x = 0, 1 \quad (136)$$

expressing the assumption that there is no diffusive flux of SO_2 into or out of the region at the boundaries. After averaging and forming the residual, (136) becomes

$$\frac{\partial u}{\partial x} = 0, \quad x = 0, 1. \quad (137)$$

The problem is now to estimate $u(x, t)$ based on the data,

$$z(x^i, t) = u(x^i, t) + v_i(t), \quad i = 1, 2, 3, 4. \quad (138)$$

Since hourly data are available, (135) can be cast into the discrete-time form (1),

$$u(k+1, x) = \mathcal{L}_x u(k, x) + w(x, t) \quad (139)$$

with $\mathcal{L}_x = 1 + a\partial^2/\partial x^2$. Observation error is estimated from the mean square error of predicted values and observed data,

$$PA(i) = \frac{1}{24} \sum_{k=1}^{24} (z_i(k) - \hat{z}_i(k/k-1))^2, \quad i = 1, 2, 3, 4. \quad (140)$$

An index of overall estimation error is

$$J = \sum_{i=1}^4 PA(i). \quad (141)$$

To apply discrete-time distributed parameter estimation theory to predict air pollution levels, we must consider three problems. The first problem is how to simulate the distributed parameter system. The second is how to determine the covariances of system and observation noise. The last is how to determine the diffusion coefficient a . For the first problem we use the Fourier expansion method and approximate the original distributed parameter system by a finite-dimensional system. For the second problem, we apply the algorithm of Sage and Husa [12] that necessitates the simultaneous application of the optimal filtering and smoothing algorithms. For the third problem we apply the maximum likelihood approach in the smoothing form [14]. We now consider these problems in more detail.

Fourier Expansion Method: It is well-known that the state $u(k, x)$ of the distributed parameter system (139) with boundary condition (137) can be represented by using the eigenfunctions $\phi_i(x)$ as follows,

$$u(k, x) = \sum_{i=1}^{\infty} u_i(k) \phi_i(x) \quad (142)$$

where

$$\begin{aligned} \mathcal{L}_x \phi_i(x) &= \lambda_i \phi_i(x), \quad x \in (0, 1) \\ \frac{\partial \phi_i(\xi)}{\partial \xi} &= 0, \quad \xi = 0, 1 \end{aligned} \quad (143)$$

and

$$\int_0^1 \phi_i(x) \phi_j(x) dx = \delta_{ij}.$$

λ_i is the eigenvalue of \mathcal{L}_x corresponding to $\phi_i(x)$. In this case, it is easily seen that the eigenfunction $\phi_i(x)$ and the eigenvalue λ_i are given by

$$\phi_1(x) = 1, \quad \phi_i(x) = \sqrt{2} \cos \pi i x, \quad i = 2, \dots$$

and

$$\lambda_i = 1 - a\pi^2(i-1)^2, \quad i = 1, 2, \dots \quad (144)$$

Then $\hat{u}(\tau, x/k)$, $p(\tau, x, y/k)$, and $A(\tau, x)$ can be represented as follows:

resented as follows:

$$\begin{aligned} \hat{u}(\tau, x/k) &= \sum_{i=1}^{\infty} \hat{u}_{ij}(\tau/k) \phi_i(x) \\ p(\tau, x, y/k) &= \sum_{i,j=1}^{\infty} \tilde{p}_{ij}(\tau/k) \phi_i(x) \phi_j'(y) \\ A(\tau, x) &= \sum_{i=1}^{\infty} a_i(\tau) \phi_i(x). \end{aligned} \quad (145)$$

Let us approximate these infinite expansions by the first N terms and define the following matrices and vectors,

$$\begin{aligned} \hat{u}(\tau/k) &= \text{Col} [\hat{u}_1(\tau/k), \dots, \hat{u}_N(\tau/k)], \\ A(\tau) &= \text{Col} [a_1(\tau), \dots, a_N(\tau)], \\ \Lambda &= \text{diag} [\lambda_1, \dots, \lambda_N], \end{aligned}$$

$$\begin{aligned} P(\tau/k) &= \begin{bmatrix} \tilde{p}_{11}(\tau/k), \dots, \tilde{p}_{1N}(\tau/k) \\ \vdots \\ \tilde{p}_{N1}(\tau/k), \dots, \tilde{p}_{NN}(\tau/k) \end{bmatrix} \\ Q(k) &= \begin{bmatrix} q_{11}(k), \dots, q_{1N}(k) \\ \vdots \\ q_{N1}(k), \dots, q_{NN}(k) \end{bmatrix} \end{aligned}$$

and

$$\Phi = \begin{bmatrix} \phi_1(x^1), \dots, \phi_N(x^1) \\ \vdots \\ \phi_1(x^m), \dots, \phi_N(x^m) \end{bmatrix}$$

where $q_{ij}(k)$ denotes the (i, j) th Fourier coefficient of $\tilde{Q}(k, x, y)$.

Then, from Theorems 3–5 we have

$$\begin{aligned} \hat{u}(k+1/k+1) &= \Lambda \hat{u}(k/k) + F(k+1) v(k+1) \\ F(k+1) &= P(k+1/k) \Phi' H'(k+1) \\ &\quad \cdot [H(k+1) \Phi P(k+1/k) \Phi' \\ &\quad \cdot H'(k+1) + R(k+1)]^{-1}, \\ P(k+1/k) &= \Lambda P(k/k) \Lambda' + Q(k), \\ P(k+1/k+1) &= (I - F(k+1) H(k+1) \Phi) \\ &\quad \cdot P(k+1/k). \end{aligned} \quad (146)$$

Furthermore, from Theorem 13 we have

$$\begin{aligned} \hat{u}(\tau+1/k) &= \hat{u}(\tau+1/\tau+1) + A(\tau+1) \Phi (\hat{u}(\tau+2/k) \\ &\quad - \hat{u}(\tau+2/\tau+1)), \\ A(\tau+1) &= P(\tau+1/\tau+1) \Lambda P^{-1}(\tau+1/\tau) \Phi^{-1}, \\ P(\tau+1/k) &= P(\tau+1/\tau+1) \\ &\quad - A(\tau+1) \Phi (P(\tau+1/k) \\ &\quad - P(\tau+1/\tau)) \Phi' A'(\tau+1). \end{aligned} \quad (147)$$

Note that the fixed-interval smoothing estimator does not depend on the matrix Φ which reflects the effect of sensor location.

Determination of the Noise Covariances: In order to determine the unknown covariance matrices of the system and observation noises, we adopt Sage and Husa's algorithm [12] given by

$$\hat{Q}(k) = \frac{1}{k} \sum_{j=1}^k (\hat{u}(j/k) - \Lambda \hat{u}(j-1/k)) \cdot (\hat{u}(j/k) - \Lambda \hat{u}(j-1/k))' \quad (148)$$

and

$$\hat{R}(k) = \frac{1}{k} \sum_{j=1}^k (z(j) - H(j)\Phi \hat{u}(j/k)) \cdot (z(j) - H(j)\Phi \hat{u}(j/k))' \quad (149)$$

where $\hat{Q}(k)$ and $\hat{R}(k)$ denote the estimated values of $Q(k)$ and $R(k)$, respectively. Note that in this identification algorithm the fixed-interval smoothing estimate $\hat{u}(j/k)$ is used.

Identification of the Unknown Parameter a : To determine the unknown parameter a we use the maximum likelihood approach in smoothing form [14]. The log-likelihood function $\gamma(k; a)$ is given from [14] by

$$\gamma(k; a) = \frac{1}{2} (\gamma_{\text{bias}} + \gamma_{\text{obs}}) \quad (150)$$

where

$$\begin{aligned} \gamma_{\text{bias}} &= -kp \ln(2\pi) - \sum_{j=1}^k \ln \det \Sigma_v(j/j-1; a) \\ \gamma_{\text{obs}} &= - \sum_{j=1}^k \{ \nu'(j; a) R^{-1} \nu(j; a) + (\hat{u}(j/k, a) \\ &\quad - \hat{u}(j-1/k, a))' Q^{-1} \\ &\quad \cdot (\hat{u}(j/k, a) - \hat{u}(j-1/k, a)) \} \end{aligned}$$

$$\nu(j; a) = z(j) - H(j)\Phi \hat{u}(j/j-1, a)$$

$$\Sigma_v(j/j-1; a) = E[\nu(j; a)\nu'(j; a)],$$

where p is the dimension of $z(k)$, and $\hat{u}(j/k-1, a)$ denotes $\hat{u}(j/k-1)$ under the condition that the unknown parameter is assumed to be a .

To maximize $\gamma(k; a)$ we use the following gradient method,

$$\begin{aligned} a_{i+1} &= a_i + G(i)\gamma_g(k; a_i) \\ \gamma_g(k; a_i) &= \left. \frac{\partial \gamma(k; a)}{\partial a} \right|_{a=a_i} \end{aligned} \quad (151)$$

where $G(i)$ is a suitable matrix. Therefore, we adopt the following recursive algorithm to identify the unknown parameters Q , R , and a :

- 1) Make an initial guess a_0 of a .
- 2) Compute $\hat{Q}(a_0)$ and $\hat{R}(a_0)$ by using (148) and (149).
- 3) Compute \hat{a}_i by using (150).
- 4) compute $\hat{Q}(\hat{a}_i)$ and $\hat{R}(\hat{a}_i)$ by using (148) and (149).
- 5) Return to three by changing i to $i+1$ and repeat until these values do not change.

TABLE I
EFFECT OF THE NUMBER OF OBSERVATION LOCATIONS ON THE OVERALL ERROR J

Number of Sensor Locations	Sensor Location	J
1	Aizumi (x^1)	39.9
	Kitajima (x^2)	83.1
	Kawauchi (x^3)	46.3
	Matsushige (x^4)	39.1
2	x^1, x^2	26.9
	x^1, x^3	25.8
	x^1, x^4	23.9
	x^2, x^3	41.3
	x^2, x^4	37.7
	x^3, x^4	29.2
3	x^1, x^2, x^3	22.0
	x^1, x^2, x^4	22.9
	x^1, x^3, x^4	23.4
	x^2, x^3, x^4	29.0
4	x^1, x^2, x^3, x^4	10.6

Numerical Results: We use the observed data from December 1–30 to identify the unknown parameter a and noise covariances Q and R . After four iterations the algorithm for determining a converged to the value $\hat{a} = 0.001$. The Fourier expansion has been truncated at $N = 4$. The estimated diagonal elements of noise covariance matrices are

$$\begin{aligned} Q_{11} &= 6.44 & R_{11} &= 0.29 \\ Q_{22} &= 1.40 & R_{22} &= 0.61 \\ Q_{33} &= 5.75 & R_{33} &= 1.96 \\ Q_{44} &= 3.68 & R_{44} &= 1.34. \end{aligned}$$

To consider the effect of the number and location of monitoring stations, we assume that we have data at only one monitoring station. In this case from the previous results of Kumar and Seinfeld [15] and Omatu *et al.* [16] we expect that the optimal sensor location is closest to the boundary. Thus, either x^1 or x^4 is the optimal single sensor location among the four monitoring stations, x^1, x^2, x^3, x^4 . In Table I we show the values of $PA(i)$ and J for several monitoring stations. We see that Aizumi or Matsushige is optimal for the one-point sensor location case. Similar conclusions hold for two or three monitoring stations. Finally, we illustrate the actual observation data and one-hour ahead predicted values for December 31 in Figs. 2–5 for Aizumi, Kitajima, Kawauchi, and Matsushige, respectively.

Comparison with Other Approaches: It is of interest to compare results of the present filtering and smoothing approaches with others available for air pollution estimation. We consider, therefore, the same SO_2 estimation problem by the following methods: 1) AR -model, 2) persistence, and 3) weighted ensemble.

The AR -model method is based on the following $AR(p)$ model

$$\begin{aligned} u_k^{(i)} &= a_1 u_{k-1}^{(i)} + a_2 u_{k-2}^{(i)} + \cdots + a_p u_{k-p}^{(i)} + e_k^{(i)}, \\ i &= 1, 2, 3, 4 \end{aligned} \quad (152)$$

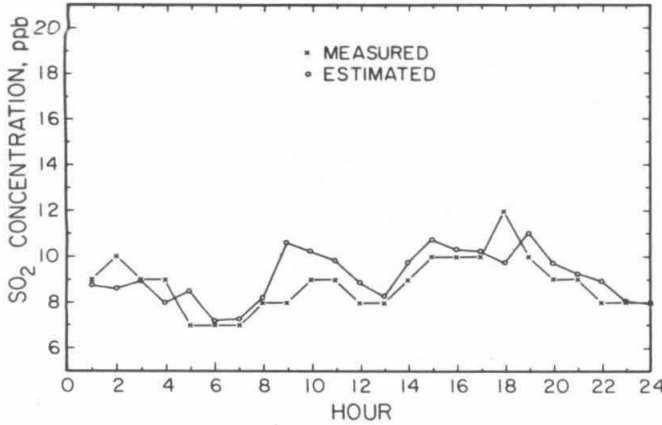


Fig. 2. Measured and estimated sulfur dioxide concentrations on December 31, 1975 at Aizumi monitoring station (x^1).

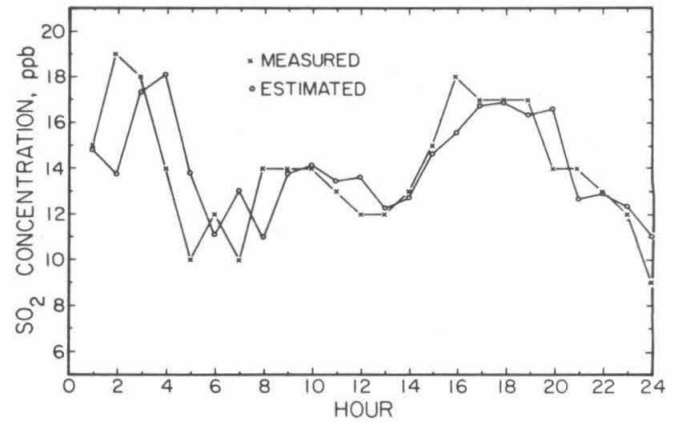


Fig. 4. Measured and estimated sulfur dioxide concentrations on December 31, 1975 at Kawauchi monitoring station (x^3).

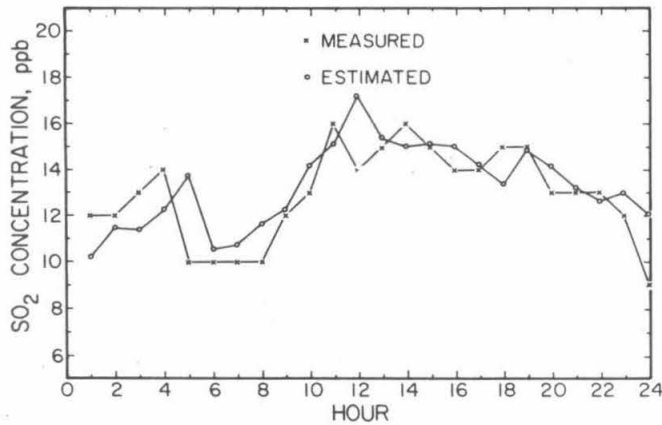


Fig. 3. Measured and estimated sulfur dioxide concentrations on December 31, 1975 at Kitajima monitoring station (x^2).

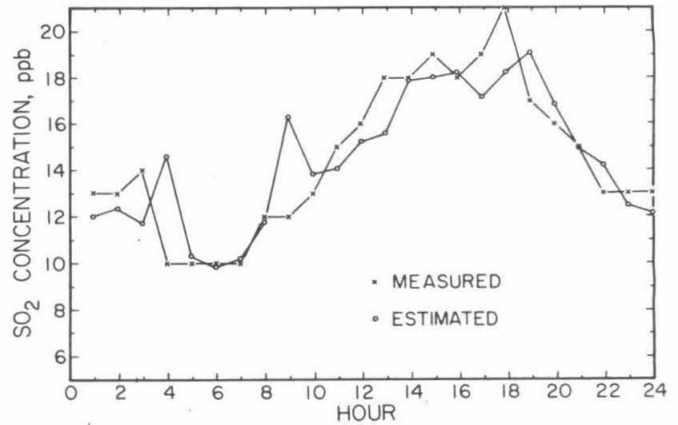


Fig. 5. Measured and estimated sulfur dioxide concentrations on December 31, 1975 at Matsushige monitoring station (x^4).

where the $u_k^{(i)}$'s are the concentration levels at time k and at monitoring station x^i , a_1, a_2, \dots, a_p are the corresponding AR -parameters, and the $e_k^{(i)}$'s are residuals. We used the Levinson algorithm to determine the AR -parameters, while the optimal order p of the AR -process is determined by using the minimum Akaike's information criterion (AIC) [20]. Then the one-hour ahead predicted concentration is given by

$$\hat{u}_{k/k-1}^{(i)} = a_1 u_{k-1}^{(i)} + \dots + a_p u_{k-p}^{(i)} \quad (153)$$

and the prediction error variance is

$$J = \sum_{i=1}^4 \left\{ \frac{1}{24} \sum_{k=1}^{24} (\hat{u}_k^{(i)} - u_{k/k-1}^{(i)})^2 \right\}. \quad (154)$$

Table II shows the AR -parameters and minimum AIC value at each monitoring station.

The persistence method consists merely of using the observation data $u_{k-1}^{(i)}$ as the one-hour ahead prediction value $\hat{u}_{k/k-1}^{(i)}$.

The weighted ensemble method uses the mean of the past observation data at each time k weighted by a linear function of the source strength as the prediction value at

time k . Based on the number of emission sources, the weighting functions are assumed here to be 0.15, 0.41, 0.26, and 0.18 at x^1, x^2, x^3, x^4 , respectively. Table III shows the performance criteria of the four methods. From Table III we can see that the present method possesses almost the same accuracy as the AR -model method. By multiplying each eigenfunction coefficient by the corresponding eigenfunction and summing them, however, the present method enables us to estimate concentrations over the entire region. Therefore, the present method is more powerful than the AR -model method.

IX. CONCLUSION

Optimal estimators for discrete-time distributed parameter systems have been derived based on Wiener-Hopf theory. A notable point of the present work is that the smoothing estimators have been derived by the same approach as the filter, thus providing a unified approach for this class of distributed parameter estimation problems. The estimation algorithms have been applied to the problem of predicting atmospheric sulfur dioxide levels in the Tokushima prefecture of Japan.

TABLE II
AR-PARAMETERS AND MINIMUM AIC (MAIC)

MAIC (Optimal p)	x^1 ($p = 5$)	x^2 ($p = 1$)	x^3 ($p = 10$)	x^4 ($p = 6$)
a_1	-0.870	-0.811	-0.721	-0.841
a_2	0.058		-0.002	-0.001
a_3	0.073		-0.074	-0.040
a_4	-0.052		0.003	0.017
a_5	-0.056		0.007	-0.079
a_6			0.016	0.133
a_7			-0.029	
a_8			0.048	
a_9			0.014	
a_{10}			0.069	

TABLE III
COMPARISON OF THE FOUR METHODS AT THE FOUR MONITORING
SITES FOR ONE-HOUR AHEAD PREDICTED VALUES-PREDICTION
ERROR SQUARED

Method	x^1	x^2	x^3	x^4	Total
Current	1.02	2.32	4.21	3.08	10.63
AR	0.99	2.36	4.31	3.11	10.77
Persistence	0.92	2.87	6.08	2.71	12.58
Weighted Ensemble	2.10	6.55	10.46	5.48	24.59

REFERENCES

- [1] R. F. Curtain, "A survey of infinite dimensional filtering," *SIAM Rev.* vol. 17, no. 3, pp. 395-411, 1975.
- [2] Y. Sawaragi, T. Soeda, and S. Omatu, "Modeling, estimation, and their applications for distributed parameter systems," in *Lecture Notes in Control and Information Sciences*, vol. 11, New York: Springer, 1978.
- [3] K. E. Bencala and J. H. Seinfeld, "Distributed parameter filtering: boundary noise and discrete observations," *Int. J. Systems Sci.*, vol. 10, no. 5, pp. 493-512, 1979.
- [4] J. S. Meditch, "A survey of data smoothing for linear and nonlinear dynamic systems," *Automatica*, vol. 9, no. 2, pp. 151-162, 1973.
- [5] S. G. Tzafestas, "Bayesian approach to distributed-parameter filtering and smoothing," *Int. J. Control*, vol. 15, no. 2, pp. 273-295, 1972.
- [6] —, "On optimal distributed-parameter filtering and fixed-interval smoothing for colored noise," *IEEE Trans. Automat. Contr.*, vol. AC-17, no. 4, pp. 448-458, 1972.
- [7] H. Nagamine, S. Omatu, and T. Soeda, "The optimal filtering problem for a discrete-time distributed parameter system," *Int. J. Systems Sci.*, vol. 10, no. 7, pp. 735-749, 1979.
- [8] T. Kailath, "An innovation approach to least-squares estimation, Part I: Linear filtering in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, no. 6, pp. 646-655, 1968.
- [9] T. Kailath and P. Frost, "An innovation approach to least-squares estimation, Part II: Linear smoothing in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, no. 6, pp. 655-660, 1968.
- [10] A. A. Desalu, L. A. Gould, and F. C. Schweppe, "Dynamic Estimation of Air Pollution," *IEEE Trans. Automat. Contr.*, vol. AC-19, no. 6, pp. 904-910, 1974.
- [11] M. Koda and J. H. Seinfeld, "Estimation of urban air pollution," *Automatica*, vol. 14, no. 6, pp. 583-595, 1978.
- [12] A. P. Sage and G. W. Husa, "Adaptive filtering with known prior statistics," *JACC*, pp. 760-769, 1969.
- [13] J. H. Seinfeld, *Air Pollution: Physical and Chemical Fundamentals*. New York: McGraw-Hill, 1975.
- [14] F. C. Schweppe, *Uncertain Dynamic Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [15] S. Kumar and J. H. Seinfeld, "Optimal location of measurements for distributed parameter estimation," *IEEE Trans. Automat. Contr.*, vol. AC-23, no. 4, pp. 690-698, 1978.
- [16] S. Omatu, S. Koide, and T. Soeda, "Optimal sensor location problem for a linear distributed parameter system," *IEEE Trans. Automat. Contr.*, vol. AC-23, no. 4, pp. 665-673, 1978.
- [17] R. M. D. Santis, R. Sacks, and L. J. Tung, "Basic optimal estimation and control problems in Hilbert space," *Math. Syst. Theory* 12, pp. 175-203, 1978.
- [18] C. N. Kelly and B. D. O. Anderson, "On the stability of fixed-lag smoothing algorithms," *J. Franklin Inst.*, vol. 291, no. 4, pp. 271-281, 1971.
- [19] S. Chirarattananon and B. D. O. Anderson, "The fixed-lag smoother as a stable finite-dimensional linear system," *Automatica*, vol. 7, pp. 657-669, 1971.
- [20] R. K. Mehra and D. G. Lainiotis, *System Identification: Advances and Case Studies*. New York: Academic, 1976, pp. 27-96.